

Análisis Inteligente de Datos: Segundo Parcial

Claudio Sebastián Castillo

05 de mayo de 2022

ANOVA

Datos

Observaciones por grupo:

Se cumplen los supuestos para su implementación?

Anova

fit del modelo

coeficientes

p-value

F-value

Plot ANOVA

Conclusión

Testear homoscedasticidad

Test de Bartlett

sensibilidad al supuesto de normalidad

Testear normalidad

Testear normalidad analizando residuos

Anova y después: post-hoc

Tukey's Honest Significant Differences (HSD)

Cuando ANOVA no funciona: test de Kruskal-Wallis

ANOVA__multivariante

Analisis Discriminante Lineal (LDA)

Datos

Explorando discriminación por pares de variable

Homogeneidad de la Varianza: Histograma VariablexGrupo

Contraste de Normalidad Univariante Shapiro-Wilk

Contraste de Normalidad MultiVariante

Outliers

Test de Royston

Test de Henze-Zirkler

Contraste de Matriz de Covarianza

Estimación de parámetros de la función de densidad ($u^*(X), E$) y cálculo de la función discriminante según aproximación de Fisher via `lda()`

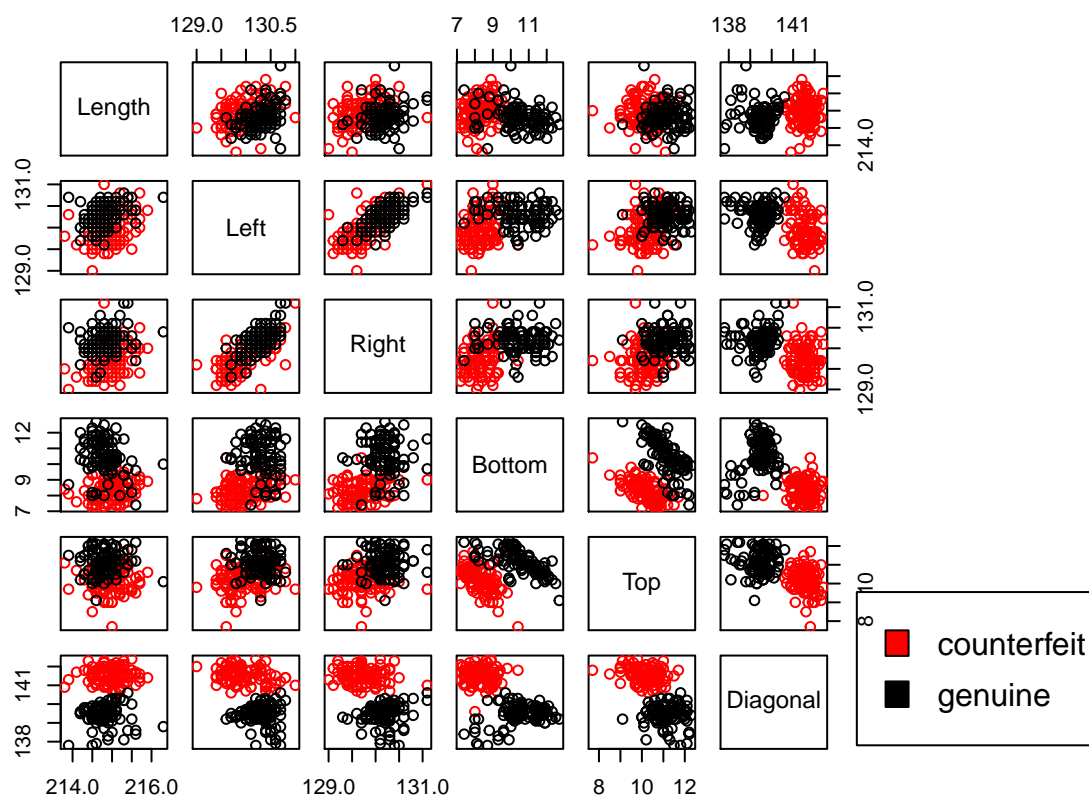
Evaluación del error: Accuracy Table

Visualización de las clasificaciones

Analisis Discriminante Cuadrático (QDA)

```
## tibble [200 x 7] (S3: tbl_df/tbl/data.frame)
##  $ Status   : Factor w/ 2 levels "counterfeit",...: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Length   : num [1:200] 215 215 215 215 215 ...
##  $ Left     : num [1:200] 131 130 130 130 130 ...
##  $ Right    : num [1:200] 131 130 130 130 130 ...
##  $ Bottom   : num [1:200] 9 8.1 8.7 7.5 10.4 9 7.9 7.2 8.2 9.2 ...
##  $ Top      : num [1:200] 9.7 9.5 9.6 10.4 7.7 10.1 9.6 10.7 11 10 ...
##  $ Diagonal: num [1:200] 141 142 142 142 142 ...
```

Explorando discriminación por pares de variable



Contraste de Normalidad Univariante Shapiro-Wilk

datosQ_tidy[["Status"]]	variable	p_value_Shapiro.test
counterfeit	Length	0.00290
counterfeit	Left	0.02372
counterfeit	Right	0.01683
counterfeit	Bottom	0.01115
counterfeit	Top	0.04909
counterfeit	Diagonal	0.00003
genuine	Length	0.25674
genuine	Left	0.00825
genuine	Right	0.01174
genuine	Bottom	0.04035
genuine	Top	0.04984
genuine	Diagonal	0.00330

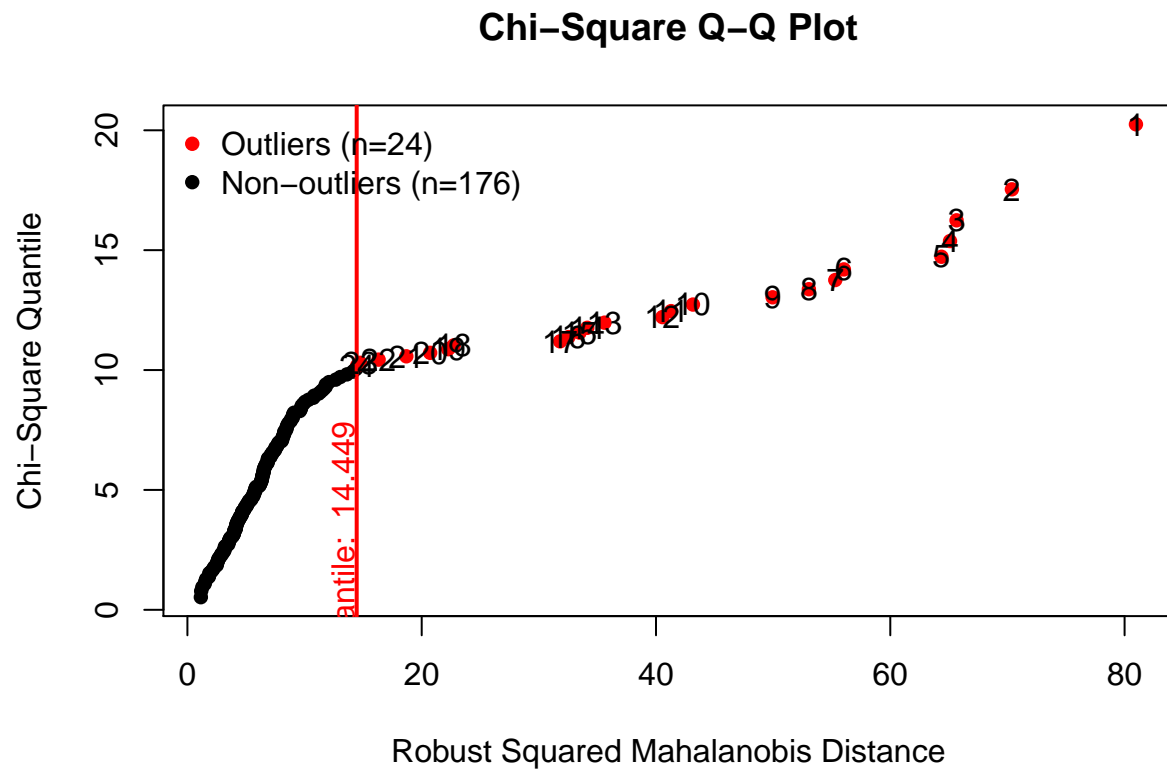
```
## [1] "H0 debe rechazarse: hay evidencia de falta de normalidad en los siguientes casos"
```

```
## # A tibble: 11 x 3
## # Groups:   datosQ_tidy[["Status"]] [2]
##   `datosQ_tidy[["Status"]]` variable p_value_Shapiro.test
##   <fct>                    <fct>          <dbl>
## 1 counterfeit             Length          0.0029
## 2 counterfeit             Left            0.0237
## 3 counterfeit             Right            0.0168
```

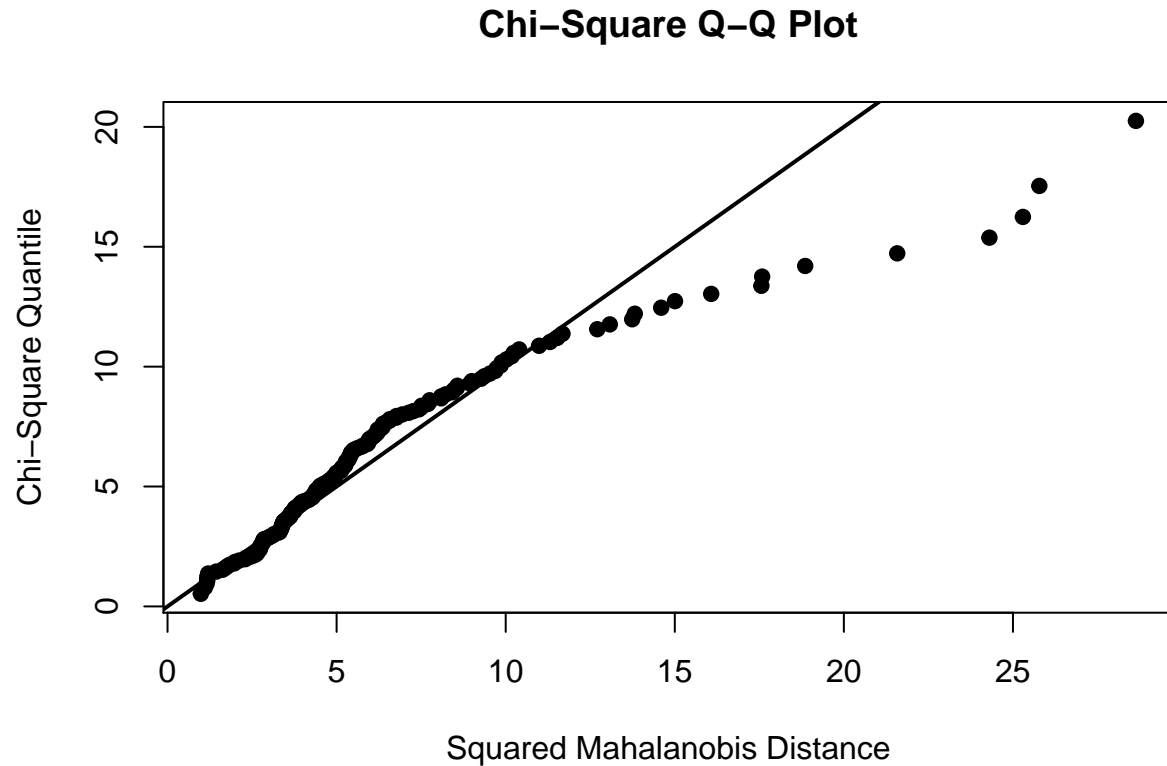
##	4	counterfeit	Bottom	0.0112
##	5	counterfeit	Top	0.0491
##	6	counterfeit	Diagonal	0.00003
##	7	genuine	Left	0.00825
##	8	genuine	Right	0.0117
##	9	genuine	Bottom	0.0403
##	10	genuine	Top	0.0498
##	11	genuine	Diagonal	0.0033

Contraste de Normalidad MultiVariante

Outliers



Test de Royston



```
##      Test      H      p value MVN
## 1 Royston 67.03927 0.0000000000005820549 NO
## [1] "H0 debe rechazarse: falta de normalidad multivariante a nivel de significancia 0.05"
```

Test de Henze-Zirkler

```
##      Test      HZ p value MVN
## 1 Henze-Zirkler 1.780591      0 NO
## [1] "H0 debe rechazarse: falta de normalidad multivariante a nivel de significancia 0.05"
```

Contraste de Matriz de Covarianza

```
##
## Box's M-test for Homogeneity of Covariance Matrices
##
## data: temp
## Chi-Sq (approx.) = 121.9, df = 21, p-value = 0.0000000000000003198
## [1] "H0 debe rechazarse: hay evidencia de que la covarianza no es igual en todos los grupos"
```

Estimación de parámetros de la función de densidad ($\hat{u}(X)$, E) y cálculo de la función discriminante según aproximación de Fisher via `qda()`

```
## Call:
```

```
## qda(temp, datosQ[[{
##   {
##     variable_factor_qda
##   }
## ]]])
##
## Prior probabilities of groups:
## counterfeit    genuine
##      0.5      0.5
##
## Group means:
##      Length    Left    Right Bottom    Top Diagonal
## counterfeit 214.823 130.300 130.193 10.530 11.133 139.450
## genuine    214.969 129.943 129.720  8.305 10.168 141.517
```

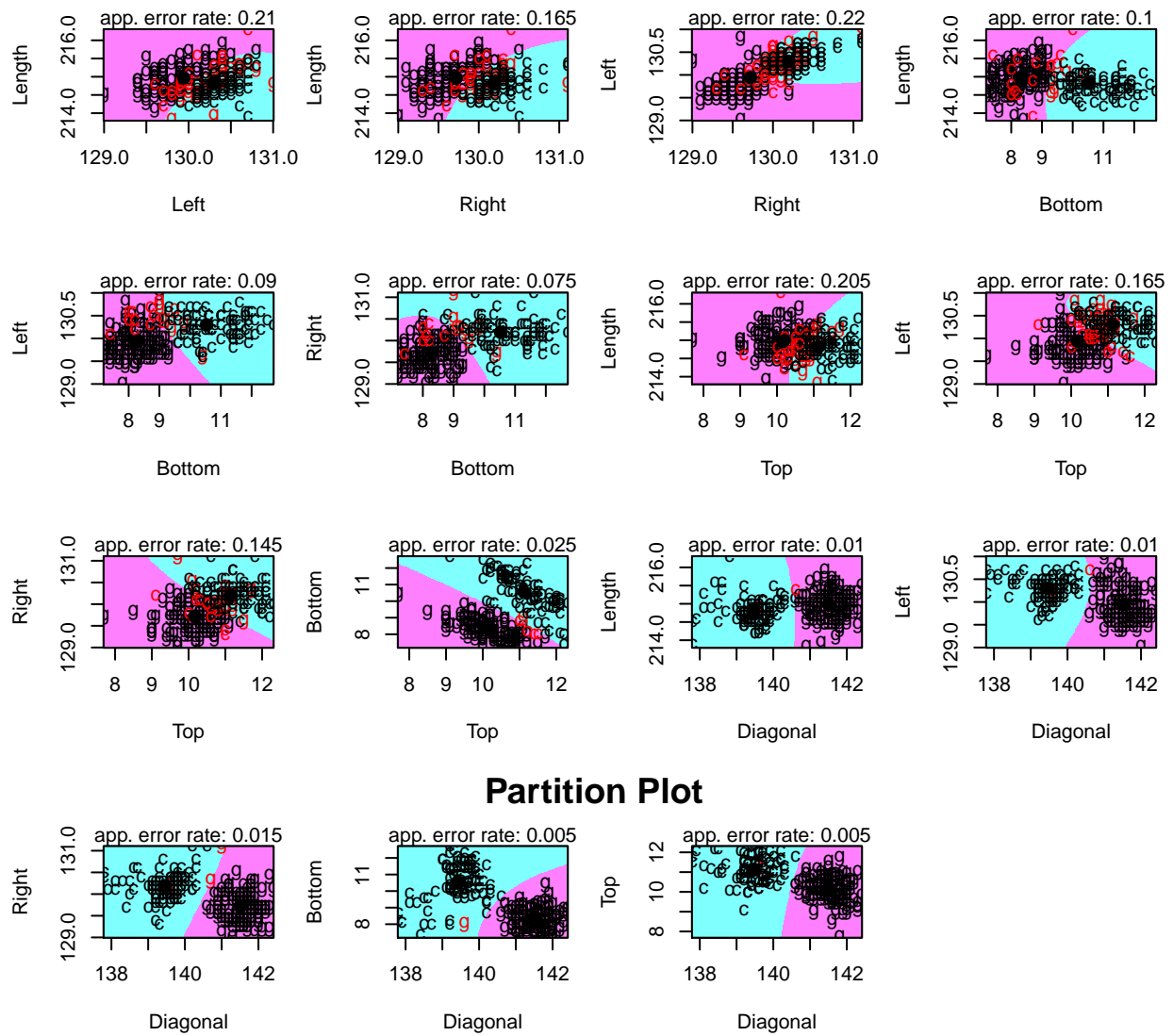
Clasificación de la nueva observación

```
## $class
## [1] genuine
## Levels: counterfeit genuine
##
## $posterior
##      counterfeit    genuine
## [1,] 0.001040119 0.9989599
```

Evaluación del error: Accuracy Table

```
##      Clase predicha
## Clase real    counterfeit genuine
## counterfeit      100      0
## genuine          1      99
## [1] "trainig_error = 0.5 %"
```

Visualización de las clasificaciones



Partition Plot